

One problem being addressed is the fact that people can supply ratings that are essentially random (due to not making the effort to provide truly meaningful ratings), or which are consciously destructive or manipulative. For instance, it has been commented that on Amazon.com, every time a new book comes out, the first ratings and reviews are from the author's friends, which are then counteracted with contradictory reviews from his enemies.

The key to solving this problem is to weight each user's ratings according to their reliability. For instance, if the author's friends and enemies are providing ratings simply to satisfy personal needs to help or hurt the author, it would be helpful if those ratings carried a lower weight than those of other users who have a past reputation for responsible, accurate ratings.

A problem solved by this invention is to provide a way to calculate that past reputation.

This reputation can be thought of as the expected "value to the system" of the user's ratings. This is bound up with the degree to which the user's ratings are representative of the real opinions of the population, particularly the population of clusters which are more appreciative of the genre into which the particular artist's work fits.

(To measure the user's overall contribution to the system, we can multiply the expected value of his ratings by the number of his ratings. Users who contribute a large number of valuable [representative] ratings are, in some embodiments, rewarded with a high profile such as presence on a list of people who are especially reliable raters.)

One can measure the representativeness of a user's ratings by calculating the correlation between those ratings and the average ratings of the larger population.

This analysis of measuring the representativeness of a user's ratings has a major limitation, however. It doesn't take into account the fact that a rating has much more value if it is the first rating on an item than if it is the 100th. The first rating will provide real guidance to those who are wondering whether to download or buy a recording before other ratings have been entered; the 100th rating will not change people's actions in a major way. So early ratings add much more

actual value to the community. Also, later raters might choose to simply copy earlier raters, so they can mislead any correlation calculations that way.

Therefore, we want to weight earlier ratings more than later ones. The question is, how much more valuable is the 1st rating than the second one, and the 2nd one more than the 3rd, etc.?

Let S be the set of all items; let N be the number of all items; for $s \in S$ and $0 < i \leq N$, s_i is the i th item. Let u be the user whose rating representativeness we wish to compute.

Let $g_{i,u}$ be the number of ratings received by s_i previous to u 's rating. (i.e., if u gives the first rating for item s_i , $g_{i,u}$ is 0.) Let t_i be the total number of ratings for the i th item.

Let $r_{i,u}$ be u 's rating of the i th item, normalized to the unit interval. Let a_i be the average of the ratings for the i th item other than u 's, also normalized to the unit interval.

Let λ_1 and λ_2 be constants.

Let q_u be the representativeness of u 's ratings, calculated as follows:

$$q_u = \frac{\sum_{i=1}^N e^{-\lambda_1 g_{i,u}} (1 - e^{-\lambda_2 t_i}) |a_i - r_{i,u}|}{\sum_{i=1}^N e^{-\lambda_1 g_{i,u}} (1 - e^{-\lambda_2 t_i})}.$$

Then q_u is a number on the unit interval which is close to 1 if the u 's ratings have tended to be predictive of those of the community as a whole, and 0 if not.

λ_1 and λ_2 are tuned for performance. λ_1 is a parameter of the cumulative exponential distribution determining the rate of "drop-off" associated with the importance of a rating as more ratings for a given item precede u 's rating. λ_2 is a parameter of the cumulative exponential distribution determining the rate at which the drop-off is associated with the number of total

ratings. For instance, if there are no ratings for an item other than u 's, the rating has no importance in calculating representativeness and is therefore given weight 0. These parameters can be set manually by intuitive understanding of the effect they have on the calculation. In some embodiments they are set by setting up a training situation in which a number of users rate the items without the means to see other people's ratings; furthermore, these users are selected and given financial or other motivation for putting the effort in to input the most accurate ratings they can generate. These controlled ratings are averaged. Then standard computer optimization techniques such as simulated annealing or genetic algorithms are used to determine values for λ_1 and λ_2 that optimize the correlation between these averages and q_u , q_u is calculated using the entire population of users in usual viewing mode (such that they could see the ratings of other users). In preferred embodiments, tuning activities are carried out within the memberships of individual clusters. That is, the controlled ratings given by members of a cluster are used to tune the parameters relative to the general ratings given by other members of the same cluster. This is carried out for each cluster. If it is deemed that there aren't enough members of some clusters to effectively tune the parameters separately for each cluster, then in such cases the values for λ_1 and λ_2 are averaged across all clusters, and clusters without enough members can use those averaged values. In addition, if a given user has created ratings in multiple clusters, some embodiments simply use the average of his representativeness numbers for all clusters as his single viewable representativeness and some clusters display separate representativeness numbers depending on the cluster in which the numbers are being viewed.

The representativeness of a user is then used for various purposes in various embodiments. In some embodiments, it is presented to artists as a reason to pay a particular user to providing ratings and reviews for new items. In further embodiments, it is used as a weight for the user's ratings when calculating overall average ratings for an item. In some embodiments, listings are provided showing the users' rankings as trustworthy raters, giving "ego gratification"; in must such embodiments these numbers are also available when viewing the user's profile, along with other information presented about the user.

It should not be construed that this invention is dependent upon the particular calculation method for representativeness which is described above.

For example, another embodiment uses the following algorithm for computing the representativeness q_u of user u :

5 Calculate the average rating for each item, not counting u 's rating. For each item, rank the population of ratings in order of their distance from the average rating. In embodiments where discrete ratings are used (that is, some small number of rating levels such as "Excellent" to "Poor" rather than a continuous scale), there will be ties. Simply give each rating a random rank to eliminate ties. For instance, if the average rating is 3, and the ratings in order of their distance from the average are, 3, 3, 4, 2, 5, 5, 1, then after randomization one of the 3's, randomly chosen, 10 will have the top rank, the other will have the next highest rank, the 4 will have the third highest rank, etc.

15 Call the distance from the average, based on these ranks, the "discrete closeness." Label the ranks such that the closest rating has rank 0, the next closest 1, etc., up to $N-1$, where N is the total number of ratings of the item. Now pick a random number on the interval $(0,1]$. Add it to the discrete closeness. Call this quantity the "real closeness" of user u to the average for the i th item and label it $p_{i,u}$. If user u 's ratings are randomly distributed with respect to the average rating for each item, then the population of $p_{i,u}$'s has a uniform distribution on the unit interval.

20 It can be shown that, due to this, the quantity $x_u = -2 \sum_{i=1}^N \log(1 - p_{i,u})$ has chi-square distribution with $2N$ degrees of freedom. A chi-square table can then be used to lookup a p-value, p'_u , relative to a given value of x_u . The quantity $p_u = 1 - p'_u$ is also a p-value and has a very useful meaning. It approaches 0 when the distance between u 's ratings and the averages are consistently close to 0, "consistently" being the key word. Also, as N increases, p_u becomes still closer to 0. It represents the confidence with which we can reject the "null hypothesis" that u 's 25 ratings do not have an unusual tendency to agree with the average of the community. So p_u is an excellent indicator of the confidence we should have that user u consistently agrees with the ultimate judgement of the community (in most embodiments, this is the community within a taste cluster).

Preferred embodiments using the chi-square approach also include weights relative to how early u was in rating each item and to take into account the number of ratings for each item. Let $w_{i,u} = e^{-\lambda_1 g_{i,u}} (1 - e^{-\lambda_2 t_i})$, where $g_{i,u}$ and t_i are defined as before. Let

$$y_u = \prod_{i=1}^N p_{i,u}^{w_{i,u}}.$$

Then

$$p'_u = \text{Prob}\{y_u \leq b\} = \sum_{i=1}^N \frac{b^{w_{i,u}}}{d_i},$$

where

$$d_i = \frac{(w_{i,u} - w_1)(w_{i,u} - w_2) \dots (w_{i,u} - w_{i-1})(w_{i,u} - w_{i+1}) \dots (w_{i,u} - w_N)}{w_{i,u}^{N-1}}$$

10 We use $p_u = 1 - p'_u$ as the measure of representativeness, with numbers closer to 0 being better, as before.

15 Finally further embodiments provide weights for one or both of the terms in the expression for $w_{i,u}$. Proper weights can be found using the same procedures as are used for finding λ_1 and λ_2 ; using genetic algorithms and other optimization techniques, in some embodiments all these weights are found at the same time.

20 In general, in various preferred embodiments of the invention, various algorithms that allow a representativeness number to be calculated which includes the predictive nature of the user's ratings are used, so the invention as a whole has no dependency on any particular method.

25 When displaying the quantities calculated as the representativeness numbers, preferred embodiments calculate rankings of the various users with respect to those numbers, or percentile rankings, or some other simplifying number, since the representativeness numbers themselves are not intuitively comprehensible to most users.

Another useful feature emerges if we take $g_{i,u}$ to be a measure of elapsed time in days between the public release of an item and the time the user rated it (which can be 0 if the review preceded

or coincided with the public release), and $\lambda_2 = \infty$. Then the approaches mentioned above for calculating representativeness can be extended to such situations as measuring the value of a user in predicting the overall long-term sales of particular items (or even to predicting stock market prices and movements and other similar applications).

5

For instance, in some embodiments, a correspondence is made between ratings and ultimate sales volumes. In one such embodiment, the following algorithm is executed. For each rating level, all items with that average rating (when rounded) are located which have been on sale for a year or longer. Then, within each cluster, average sales volumes for each rating level's items are calculated. Then this correspondence is used to assign "sales ratings" to each item based on the total sales of that particular item; the actual sales are matched to the closest of the rating-associated levels of average sales, and the corresponding rating is used as the sales rating. (If there hasn't yet been enough activity in a particular cluster to conduct this exercise meaningfully, system-wide averages are used.)

10

In this embodiment $p_{i,u}$ is computed using rankings of distances from the sales rating rather than from the average rating. Then λ_2 is set to ∞ (in other words, the $(1 - e^{-\lambda_2 t_i})$ term is set to 1). Then we calculate the representativeness, p_u , as before.

15

As with the case of calculating representativeness with respect to general ratings, it should not be construed that this invention is dependent upon the specific calculations given here for calculating a user's ratings' representativeness with respect to sales; other calculations which accept equivalent information, including the user's ratings, the sales volumes, and time data for ratings and sales (or, equivalently, elapsed time data), outputting a representativeness which involves a predictive component, will also serve the purpose of providing equivalent means for use by the invention overall.

20

25

For instance, in some embodiments, a rank-based technique is used for calculating representativeness. In one such embodiment, time data is used to determine the items that the user rated soon after their release (or at or before their release) and that have now been on the market long enough to meaningfully measure sales volumes. These items are used to perform

30